

Proszę wpisywać odpowiedzi w ramkach

Zadanie 1 (3 pkt)

Badamy podobieństwo globalne algorytmem Needlemana-Wunscha dla sekwencji BAABB i ABBA. Stosujemy liniową karę za przerwę $d = -2$. Macierz podobieństwa pokazano obok, n jest liczbą liter Twojego nazwiska podzieloną na 3, dzielenie całkowite, bez reszty.

	A	B
A	$3 * n$	$-n$
B	$-n$	$2 * n$

Uzupełnij macierz podobieństwa, macierz pomocniczą. Podaj najlepsze dopasowanie, jeżeli jest kilka, podaj jedno z nich.

		B	A	A	B	B
A						
B						
B						
A						

	A	B
A		
B		

Rozwiązanie:

Zadanie 2 (3 pkt)

Sekwencje TAAA, TATA, AAATTTT, ATTAA i TTTAT użyto do asemblacji pod-grafem de Bruijna (bez wag w krawędziach) 4 rzędu (w wierzchołkach są sekwencje o długości 3 symboli).

Narysuj graf.

Odpowiedz na pytanie dotyczące wyniku asemblacji, wpisując a, b albo c:

- a) wynik nie istnieje,
- b) istnieje dokładnie jedna sekwencja wyjściowa,
- c) istnieje więcej niż jedna sekwencja wyjściowa.

Jeżeli istnieje sekwencja wyjściowa (jedna lub kilka) jest nią:

Zadanie 3 (3 pkt)

Używamy metody grupowania parami ze średnią arytmetyczną (UPGMA) dla 5 taksonów (A, B, C, D, E). Macierz odległości podano obok.

	B	C	D	E
A	2	3	5	5
B		5	3	7
C			2	5
D				7

Narysuj drzewo filogenetyczne.

Zadanie 4 (3 pkt)

Posługujemy się monetami A i B, obserwując sekwencje rzutów *RRRO* (orły i reszki). Podaj najbardziej prawdopodobną sekwencję użytych monet, zakładając, że to doświadczenie może być opisane ukrytym modelem Markowa pokazanym niżej.

$$Q = \{A, B\},$$

$$V = \{O, R\},$$

$$P_A = 1, P_B = 0.$$

	A	B
A	$\frac{3}{5}$	$\frac{2}{5}$
B	$\frac{1}{5}$	$\frac{4}{5}$

	O	R
A	$\frac{1}{2}$	$\frac{1}{2}$
B	0	1

Uzupełnij macierz pomocniczą w algorytmie Viterbiego

	R	R	R	O
A	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
B	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Podaj najbardziej prawdopodobną sekwencję użytych monet:

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------

Pytanie (2 pkt)

Jakie zagadnienie / zagadnienia warto byłoby dodać do MBI

Notatki lub uwagi do Roberta Nowaka:

Zadanie 5 (3 pkt)

Poniżej podano fragment pliku SAM (kolumny POS CIGAR SEQ).
1 6M tatact
2 6M atacta
3 6M tattag
4 6M attagc

oraz odpowiadający fragment sekwencji kodującej

```
>chrZ:1-15T  
atactagcttgcat
```

Opis standardu CIGAR oraz tablica kodonów jest na kolejnej stronie.

1. Określ sekwencję aminokwasów, które są kodowane przez tę sekwencję zakładając, że na pozycji 1 zaczyna się nowy kodon

2. Czy dane z pliku SAM pozwalają wnioskować o wystąpieniu wariantu(ów)?

Jeżeli tak:

- a) określ typ wariantu(ów) (SNV / Indel ?),
- b) jaki jest jego (ich) efekt na białko,
- c) czy jest (są) to wariant(y) hetero czy homo-zygotyczny(e)?

3. Zaproponuj modyfikację pliku SAM tak aby znalazły się dowody na wystąpienie jedno-nukleotydowej delecji heterozygotycznej.

Notatki lub uwagi do Tomasza Gambina:

6. CIGAR: CIGAR string. The CIGAR operations are given in the following table (set '*' if unavailable):

Op	BAM	Description	Consumes query	Consumes reference
M	0	alignment match (can be a sequence match or mismatch)	yes	yes
I	1	insertion to the reference	yes	no
D	2	deletion from the reference	no	yes
N	3	skipped region from the reference	no	yes
S	4	soft clipping (clipped sequences present in SEQ)	yes	no
H	5	hard clipping (clipped sequences NOT present in SEQ)	no	no
P	6	padding (silent deletion from padded reference)	no	no
=	7	sequence match	yes	yes
X	8	sequence mismatch	yes	yes

- “Consumes query” and “consumes reference” indicate whether the CIGAR operation causes the alignment to step along the query sequence and the reference sequence respectively.
- H can only be present as the first and/or last operation.
- S may only have H operations between them and the ends of the CIGAR string.
- For mRNA-to-genome alignment, an N operation represents an intron. For other types of alignments, the interpretation of N is not defined.
- Sum of lengths of the M/I/S/=/X operations shall equal the length of SEQ.

Tablica kodonów:

		Second base					
		U	C	A	G		
First base	U	UUU } Phenylalanine F UUC } UUA } Leucine L UUG }	UCU } Serine S UCC } UCA } UCG }	UAU } Tyrosine Y UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine C UGC } UGA } Stop codon UGG } Tryptophan W	U	C
	C	CUU } Leucine L CUC } CUA } CUG }	CCU } Proline P CCC } CCA } CCG }	CAU } Histidine H CAC } CAA } Glutamine Q CAG }	CGU } Arginine R CGC } CGA } CGG }	C	A
	A	AUU } Isoleucine I AUC } AUA } AUG } Methionine start codon M	ACU } Threonine T ACC } ACA } ACG }	AAU } Asparagine N AAC } AAA } Lysine K AAG }	AGU } Serine S AGC } AGA } Arginine R AGG }	A	G
	G	GUU } Valine V GUC } GUA } GUG }	GCU } Alanine A GCC } GCA } GCG }	GAU } Aspartic acid D GAC } GAA } Glutamic acid E GAG }	GGU } Glycine G GGC } GGA } GGG }	G	U
						Third base	
		U	C	A	G	U	C
		C	A	G	U	C	A
		A	G	U	C	A	G
		G	U	C	A	G	U